Designing a Guardian Angel: Giving an Automated Vehicle the Possibility to Override its Driver

Steffen Maurer[1]
Rainer Erbach[1]
Issam Kraiem[2]
Robert Bosch GmbH
Renningen, Germany
[1]firstname.lastname@de.bosch.com
[2]Issam.Kraiem@hs-pforzheim.de

Susanne Kuhnert[3]
Petra Grimm[3]
Institute of Digital Ethics
Hochschule der Medien
Stuttgart, Germany
[3]lastname@hdm-stuttgart.de

Enrico Rukzio
Institute of Mediainformatics
Ulm University
Ulm, Germany
enrico.rukzio@uni-ulm.de

ABSTRACT

A function of an automated driving vehicle that can override a human driver while driving manually could work as a guardian angel in the car. It can take over control if it detects an imminent accident and has a possibility to avoid it. Because of the urgency of the intervention, there is not enough time to warn the driver in advance. In a study, feedback was collected from users how they perceived such an action while driving in a simulator. Additional feedback was collected about the general design and user interface of such a system. From an ethical point of view, we discovered discrepancies in the views of our participants regarding automated driving functions that need to be addressed in future development.

Author Keywords

Cooperative driving; ethics in human-computer-interaction; overriding the driver; guardian angel; ethics-by-design; user study

CCS Concepts

• **Human-centered computing~Interaction design theory, concepts and paradigms** • Human-centered computing~User studies • Human-centered computing~Auditory feedback • *Security and privacy~Human and societal aspects of security and privacy*

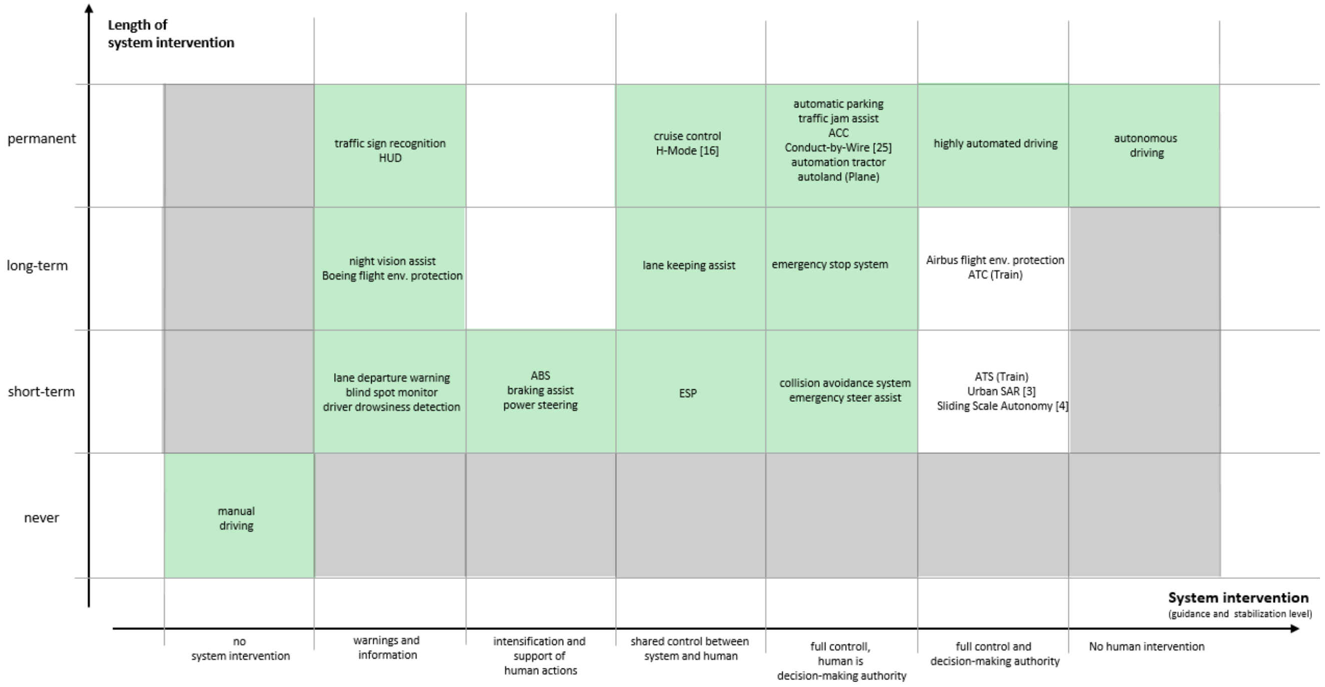
INTRODUCTION

Automated and autonomous driving is currently being developed throughout the world. Not only the traditional car companies are trying to bring self-driving cars onto the roads as quickly as possible, but also companies that up to now had nothing to do with the development of cars [32]. The current question of interest is not any longer, if automated driving will be reality, but when it will be available. Until the upcoming automated vehicles will reach SAE level 5 [35] and are able to handle all aspects of an entire journey on their own, the human driver is still needed to handle at least parts of the journey. Self-driving cars could greatly reduce injuries in traffic, as most of the accidents are caused by human error [24]. As long as the driver is needed to perform actions, at least from time to time, this risk will not decrease. Even though the automation is not performing the driving task during manual drive, the sensors of the (automated) car are still active and can sense the surrounding environment. This can be used for passive comfort functions like registering parking spots [30], but also to detect risky situations and possible accidents. If the driver or another road user makes a mistake, the car could sense this. There might be situations where the car can determine a possibility to avoid an accident and prevent car body damage or even injuries. In the event of the car sensing a threat and being able to determine a strategy to avoid it, it is ethically obliged to intervene. Such a situation can arise quickly, there could not be enough time to inform the driver of the upcoming situation and instruct him or her how to avoid it. In this case, the automation has to take over control, impeaching the driver until the hazardous situation is resolved and the car is back in a safe state. Trust in the automation to be able to make correct decisions and to predict traffic behavior is already present today, otherwise automated driving would not be possible at all and companies would not advertise automated driving to be commercially launched within the next years.

RELATED WORK

In current vehicles, a huge number of assistive systems can be found and the number of available systems is still increasing [26]. The systems can be categorized in two major categories: safety-oriented functions and comfort-oriented functions. While anti-lock braking systems (ABS) clearly fall in the first category, systems like adaptive cruise control (ACC) are mainly comfort-oriented, although they provide a safety aspect, too. The main difference between the two categories are that safety-related functions are always active, whereas comfort-related functions can be turned on and off by the driver.

AutomotiveUI '18, September 23–25, 2018, Toronto, ON, Canada
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5946-7/18/09…$15.00
https://doi.org/10.1145/3239060.3239078



Length of system intervention	no system intervention	warnings and information	intensification and support of human actions	shared control between system and human	full controll, human is decision-making authority	full control and decision-making authority	No human intervention
permanent		traffic sign recognition HUD		cruise control H-Mode [16]	automatic parking traffic jam assist ACC Conduct-by-Wire [25] automation tractor autoland (Plane)	highly automated driving	autonomous driving
long-term		night vision assist Boeing flight env. protection		lane keeping assist	emergency stop system	Airbus flight env. protection ATC (Train)	
short-term		lane departure warning blind spot monitor driver drowsiness detection	ABS braking assist power steering	ESP	collision avoidance system emergency steer assist	ATS (Train) Urban SAR [3] Sliding Scale Autonomy [4]	
never	manual driving						

Figure 1: Taxonomy of different assistive systems in different transportation modes. Green highlighting indicates that at least one automotive technology is present in the respective category. [17]

Feedback Methods

A very important part of developing new ADAS is to design the feedback method. The assistive system needs to provide information for the driver about its activity. Current systems usually communicate with the driver through "*visual and auditory display modalities*"[28]. The more advanced the ADAS gets, the more do "*car makers need to attend not only to the design of autonomous actions but also to the right way to explain these actions to the drivers*" [15]. Facilitating trust of the driver in the system is the key factor and Nothdurft et al. showed that giving an explanation that justifies and transparently explains why the action was happening "*are the most promising ones for incomprehensible situations in HCI*" [19]. Not only is the type of explanation a factor that needs to be considered, but also the time when it is happening [15]. A pro-active explanation that is told every time before an automated overtaking maneuver might become annoying for the passengers very fast [19].

Overwriting the driver

ADAS cannot only be classified by categorizing them into comfort- and safety-related functions or by how and when feedback is provided. Another classification approach is to use the level of system intervention and the duration of the current intervention. In Figure 1 such classification is shown. The x-Axis classifies the system intervention level. The categories used are based on the levels of automation from Sheridan and Verplank [22]. The y-Axis classifies the duration of the respective system intervention. Categories with systems from the automotive range of use are marked in green, categories that would not contain any useful systems are marked in grey. The taxonomy shows that there are categories that do contain systems but not from the automotive context [17]. These systems are capable of overriding the driver/operator for a certain amount of time. In critical situations they can decide on their own, even against the human. Such "*hard automation*" [27] can be found in Airbus' planes or in trains. Similar system capabilities were successfully included in robot behavior [3,4]. In the coming age of highly automated driving, such systems might be more useful than ever. With the increasing possibility to hand over driving tasks to the automation, drivers might face a decrease in their abilities to safely operate the vehicle at all times [1], as already observable with airline pilots [29].

GUARDIAN ANGEL IN THE CAR

In the publicly funded German project "KoFFI" (stands for "Cooperative driver-vehicle-interaction") [34] the driver and the car are becoming equal partners. One of the ideas of this project is to set the car (hierarchically) above the driver in safety critical situations. For example during situations with low viewing distance (fog, darkness, etc.) the car might have a better understanding of a situation due to its sensors, as these are not limited by sight (radar, etc.). If the driver is driving manually and does make a fatal mistake that could be easily corrected by the car, the automation should take over and prevent any accident. The automation can even take over control if the driver is unfit to drive, distracted or falls asleep. A "*redistribution of autonomy*" would be the result as formulated in a partnership model by Both and Weber [2]. Such a system can work as a guardian angel that is accompanying the driver on his or her journeys, intervening in critical situations. There are many open questions how to design the interaction between the guardian-angel system and the driver in a way that the driver embraces the system's intervention.

Ethical guidelines

The task of ethics in general is to give guidelines [20] that are universally valid and at the same time practically realizable. They have to give tools and criteria "*with which planned or ongoing research can be assessed with regard to possible ethically relevant conflicts*" [20, translation by authors]. Ethical requirements for automated and connected driving, like the ones provided by the German government in June 2017 [31], shall be systematically extended by the research in the KoFFI-project. The Insitute of Digital Ethics (IDE) [33] of the "Hochschule der Medien" in Stuttgart is responsible for questions regarding acceptance and trust in the human-machine-interaction within the project KoFFI. An ethics-by-design approach makes sure to sensitize all project partners in all phases of research. Ethical design standards can be derived by a deliberate reflection from the results of research. The goal is to create a humane design. The IDE is using empirical data to provide and justify detailed and realistic requirements for the engineers and designers. This is called experimental philosophy [18,7]. A method used in experimental philosophy is narrative research [6] which uses narrative elements like the ones found during the design process of a system, especially in use cases and scenarios [5,14] as they transfer the values and views of the designer and engineers [23].

Research Questions

The first question that needs an answer is: *Are drivers open to a system capable of overriding them and do they want to have one in their car? (Q1).* Related to that is the situation in which the system should intervene: *Do the drivers only allow an action of the system if the situation is critical or also in uncritical situations? (Q2)*. The next question summarizes all related interaction and interface design decisions: *How does a "guardian angel" need to communicate its intervention? (Q3)* Lastly, if the endangerment is over, the controls of the car must be shifted back to the human driver. To do this in a safe manner it is important to know what people are doing during an override situation: *How do drivers behave while the system is actively overriding them? (Q4)* A system that can decide on its own to take control from the driver affects the drivers self-determination. This is an interesting "*ethically relevant conflict*" [20] and raises the question: *Do people already have a consistent mental concept of the role model of a driver of an automated vehicle*? (Q5) This is important, as an uncertainty in the task of the human could lead to possible operating errors of the drivers.

USER STUDY

Both research questions Q1 and Q2 can be answered by simply asking drivers about their opinion. Nevertheless, it is important that drivers have a good idea what an overriding functionality works and feels like. Not all drivers know intervening ADAS like the emergency braking assist. Few to no drivers have experience with a car that can drive on its own. To give all participants the same idea what such a system could work like all participants had to experience an example system in action. This is why we conducted a user study in a driving simulator. The other major goal of the user study was to gather (qualitative) feedback from the participants how it feels to be overridden by the car while driving manually. This was done under two major conditions, a safety-related override and a comfort-related override. In order to be able to test this in a safe environment, we used a driving simulator.

Participants

In accordance with Hwang and Salvendy's 10±2 rule [13], 24 participants were recruited from the Robert Bosch GmbH and from the Pforzheim University of Applied Sciences. The only requirement for the experiment was for the participants to have a valid driver's license. 12 of the subjects were male and 12 female. 10 participants were in the age group 20-29, 10 in the group 30-39, 3 in the group 40-49 and 1 in the group 50-59. The number of years the participants have had their driver licenses ranged from 1 to 37 years, with a mean of 12.2 years (SD = 9.45). Except for two subjects all stated to use a car on at least a weekly basis, with the two stating to use a car less than once a month. 10 of the participants have one or more driver assistive systems in their car, 6 have at least experienced such a system in another person's car and 8 participants had no experience with an assistive system at all. To see if the participants are already familiar with speech-based systems that can provide detailed answers, the subjects were asked about their experience with digital assistant technologies: 10 of the participants used or are using a system like Alexa or Siri, while the other 14 participants stated to not using one. All but one participant attributed themselves a high or very high affinity for technology.

Driving Simulator

The driving simulator we used in the study is located at Bosch in Renningen, Germany. It consists of a cockpit from a BMW 3-series, which is mounted on DBOX-actuators. In front of the shell, three 4k monitors are positioned. Behind the cockpit three smaller HD monitors are installed. While the front monitors show the simulated road ahead, the rear monitors are used to display parts of the view behind the

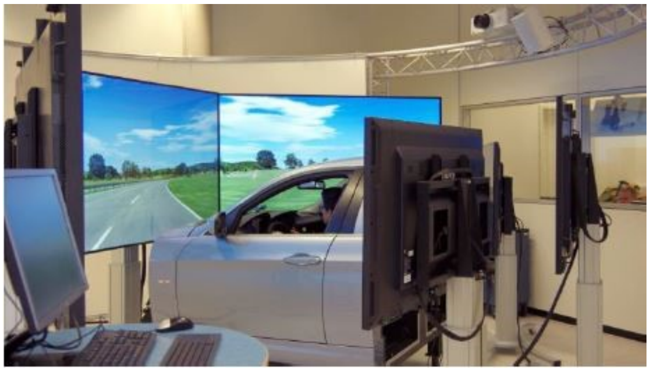


Figure 2: The driving simulator setup used in the user study

car, to allow the driver to use the car's mirrors as he or she is used to in a real car. As simulator software, SILAB [36] is used in version 5.1. A separate room connected with windows is available as an operator control room and provides access to all relevant simulator functions via two connected operator PCs. The simulator was fitted with two GoPro cameras, one mounted behind the middle mirror and facing the participant, to record the facial reactions and the steering wheel interactions. The other camera was mounted behind the pedals and recorded the feet of the participant and how he or she used the car's pedals during driving. This was done to not only collect data whether or not a certain pedal was pressed, but to gather insights of foot movements not detectable by the simulator's log files. If, for example, a participant has his or her foot readily on the pedal but does not yet press it, this would not be detectable if there was no camera used to collect the data. This can help to answer Q4 as the behavior of the driver is directly observable.

Procedure

At the beginning of the study, the experimenter greeted the participants and asked them to sign a declaration of consent. After that, a demographic questionnaire was handed to the participant to gain statistical data of age, sex, car usage and previous knowledge of (driver) assistive technology.

Then the participant was taken to the driving simulator and asked to set the driving seat and mirrors to their comfort. To familiarize the subject with the simulated environment, he or she was presented with three test tracks. The first test track consisted of a flat country road where accelerating, steering, and driving at a certain speed and braking was introduced and practiced. The second track was a straight test track where traffic signs indicated positions at which the participant had to accelerate or decelerate to a given speed. This track should help estimating distances and speeds in the simulator. The last track was a set of junctions where the participant had to turn left and right in order to get used to the feeling of turning in the simulator.

Afterwards the participant was taken to a poster on the wall and informed that he or she will be testing a new driver assistive system called "KoFFI". The experimenter explained the characteristics of the KoFFI-system, such as the ability to intervene in hazardous situations. Next part of the study were two different routes in the simulator. The order of the two conditions were counterbalanced with the participants, regarding sex, age and driving experience.

Route 1 was used to test a driver override in a safety-critical condition. The route consisted of six T-junctions linked with tracks from 1 km to 3 km in length. The participants were instructed to drive to the town of Renningen and at the intersections the drivers had to turn either right or left, indicated by street signs showing the way to Renningen. To avoid disruptions in case of a wrong turn the subjects had to drive a 2 km long detour track, until they reached the right track again. After the track, a final T-junction was reached, where a police car would cross with high speed as soon as the participants drove off at the stop line. The automation braked automatically and steered a bit to the left to avoid a collision with the other car. Feedback for the driver was provided to one half of the participants with a beeping sound, consisting of a "double-beep", repeated once, like the one used in current systems, for example in an emergency braking system. The other half was given a spoken feedback of KoFFI, where the automation stated that it had to override the driver to prevent an accident. Both auditory feedbacks were played as soon as the system started braking.

Route 2 was used to test a driver override as a non-safety critical condition. This was thought to be some kind of comfort function to prevent the driver from driving a detour. The participant had to drive a route to Renningen again, consisting of six junctions with varying tracks lengths in between. The main difference to route 1 was that the subjects did not have to turn left or right at any of these junctions. On the track to the final intersection, the participants were asked to use their smartphone. They had to take a picture of the (simulated) landscape around them and write a message to a friend. This was done to make the driver look away from the road and therefore miss the final traffic sign, indicating to turn left at the final junction. Due to that, all 24 participants were not prepared to turn left at the last intersection and missed the turning lane. At the last possible moment, the automation was braking hard and turning left. Directly behind the junction we placed a town sign of Renningen to show the driver that this is the right way. Again, the automation provided feedback either with a beeping sound or with a spoken explanation.

Participants were recorded by the afore mentioned GoPro cameras while driving the two test routes, but not during the simulator familiarization. As already mentioned, both routes and the feedback-condition were counterbalanced. Directly after each route the participant had to exit the simulator and answer three questionnaires, regarding the experienced override situation. First, a NASA TLX questionnaire [9] in the raw version [10] and an AttrakDiff questionnaire [11] had to be answered. After that, a custom questionnaire was handed to the participants, where they were asked if they think a system that is able to perform an override while driving, makes sense to them. We also asked whether the participant would like to have such a system included in their car and to explain why or why not. In addition, we questioned the experienced way of feedback and in case the participant disliked it, he or she was asked to describe their preferred feedback method.

To answer ethical research questions, we handed the participants another questionnaire at the end of the study. It consisted of five questions regarding ethical aspects as autonomy, responsibility and trust in the case of automated driving. The questions were open-ended and as it was part of a narrative research approach, the participants were instructed to write down anything that comes to their mind.

Results: NASA RTLX

To help answer Q3 it is important to know if there are differences in the use of different communication methods. One value that has to be taken into account is the task-load a person experiences while using a certain system. In the NASA task-load-index questionnaire the participants have to rate six scales on a 100-points range, where 0 marks a very low demand in the respective category and 100 marks a very high demand [9]. These ratings are then combined to the task-load-index. In the safety-group the raw task-load-indices for beeping sound feedback is higher ($M_{Beep} = 39.0$, $SD_{Beep} = 16.8$ and $M_{Voice} = 37.8$, $SD_{Voice} = 20.3$) than voice-feedback. The same is true for the comfort scenario ($M_{Beep} = 58.3$, $SD_{Beep} = 19.3$ and $M_{Voice} = 52.3$, $SD_{Voice} = 25.6$). An independent samples t-test revealed no significant differences between the feedback methods within the respective scenario groups (df = 22; $t_{Safety} = 0.27$, $p_{Safety} = 0.79$ and $t_{Comfort} = 0.66$, $p_{Comfort} = 0.51$).

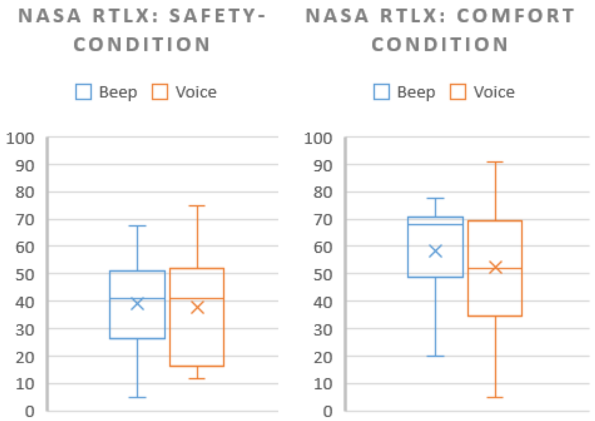
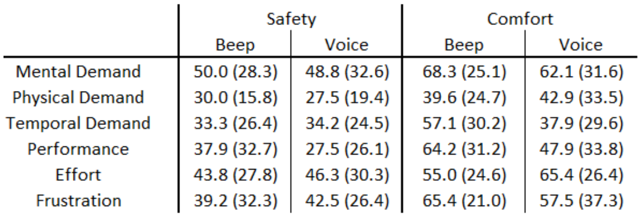


Figure 3: Results of the raw NASA TLX questionnaire used in our user study.

The comfort scenario constantly had a higher task-load-index, which presumably originates in the situation where the override happened. The participants had to write a message on their smartphone and trying to drive when the automation override happened, while in the safety-scenario they could focus completely on driving the car. Another interesting observation was the high difference between the highest and lowest task-load, especially in the comfort szenario with spoken feedback where one participant had a task-load of 5, while another wone had a task-load of 90.



	Safety		Comfort	
	Beep	Voice	Beep	Voice
Mental Demand	50.0 (28.3)	48.8 (32.6)	68.3 (25.1)	62.1 (31.6)
Physical Demand	30.0 (15.8)	27.5 (19.4)	39.6 (24.7)	42.9 (33.5)
Temporal Demand	33.3 (26.4)	34.2 (24.5)	57.1 (30.2)	37.9 (29.6)
Performance	37.9 (32.7)	27.5 (26.1)	64.2 (31.2)	47.9 (33.8)
Effort	43.8 (27.8)	46.3 (30.3)	55.0 (24.6)	65.4 (26.4)
Frustration	39.2 (32.3)	42.5 (26.4)	65.4 (21.0)	57.5 (37.3)

Figure 4: Results of each subscale of the TLX questionnaire, mean value with standard deviation in brackets

Regarding the mean values of the respective parts of the TLX (mental demand, physical demand, temporal demand, performance, effort and frustration) there are again no statistical significant differences. Again, only the beep and the voice condition were compared for each respective subscale in the two main scenarios with an independent samples t-test. (df = 22; $t_{S\_mental} = 0.16$, $t_{S\_physical} = 0.43$, $t_{S\_temporal} = -0.10$, $t_{S\_performance} = 0.84$, $t_{S\_effort} = -0.26$, $t_{S\_frustration} = -0.25$, $t_{C\_mental} = 0.52$, $t_{C\_physical} = -0.25$, $t_{C\_temporal} = 1.66$, $t_{C\_performance} = 1.30$, $t_{C\_effort} = -0.98$, $t_{C\_frustration} = 0.65$, all p-values > 0.1). Yet, there are some interesting observations: The performance value is in both scenarios lower if the automation gave a spoken explanation of the override. The lower the value in the performance category is, the better the participant rated his or her achieved outcome. That means, the participants with the spoken feedback had the feeling "they did better", in contrast to when there was only a beeping sound. However, both voice groups seemed to need more effort to achieve their level of performance than the respective beeping groups. While the use of voice feedback could lower the values for frustration and temporal demand in the comfort scenario, there was a contrary effect in the safety scenario, despite being quite low. Those results (effort needed and perceived performance) factor into the acceptance of a system by the users and therefore help to answer Q1.

Results: AttrakDiff

Another factor that should be examined to answer Q1, Q2 and Q3 is the perceived effectiveness and attractiveness of the system. The AttrakDiff questionnaire measures these two main dimensions of a product, called hedonic and pragmatic quality. The first one is an expression of how much the user wants to own the tested product and the second one is how good the product is designed to solve the specific task. A good and desired product has high values in both categories [11]. The questionnaire consists of 28 semantic differentials with seven gradations. Hassenzahl et al. developed AttrakDiff with the ability to divide the hedonic quality into two groups, identity (identification with the product) and stimulation (stimulating the senses of the user). The results for the study conditions split in the four dimensions of AttrakDiff (pragmatic quality PQ, hedonic quality – identity HQ-I, hedonic quality – stimulation HQ-S and attractiveness ATT [11]) is shown in figure 5. All but the comfort with beeping sound condition receive positive values for the pragmatic quality and the hedonic-identity dimension. In both tested scenarios, safety and comfort, the participants that experienced the voice feedback attributed a higher hedonic and a higher pragmatic quality to the system. The overall attractiveness of the tested conditions clearly shows a favor of the safety related system and spoken feedback.

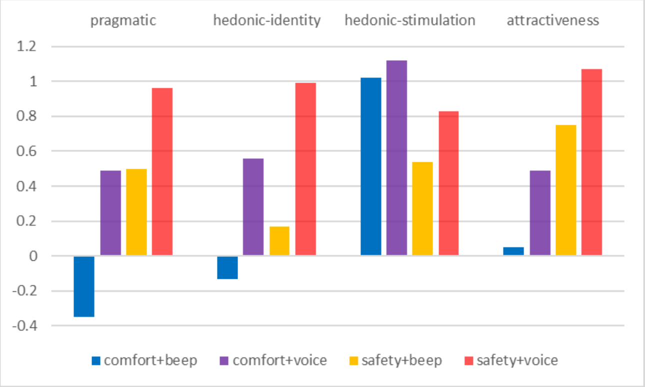
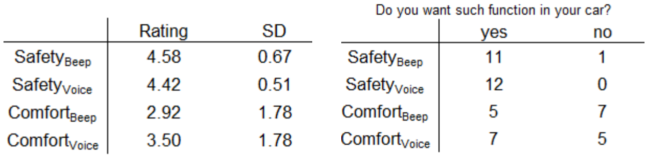


Figure 5: In-detail results of our AttrakDiff questionnaire.

Results: Qualitative Feedback

To help answering Q1, Q2 and Q3, we asked the participants to state their opinions on the usefulness of an in-car system that can overwrite their actions, to prevent an accident or to avoid minor driving mistakes such as missing a turn. The scale ranged from 0 (not at all useful) to 5 (highly useful). The results for each respective testing condition are shown in figure 6. Directly afterwards the participants were asked to state whether or not they would like to have such a function in their car and why or why not. The answers were categorized into "yes" or "no" and the number of answers in each category were counted (see figure 6).



	Rating	SD
$Safety_{Beep}$	4.58	0.67
$Safety_{Voice}$	4.42	0.51
$Comfort_{Beep}$	2.92	1.78
$Comfort_{Voice}$	3.50	1.78

Do you want such function in your car?

	yes	no
$Safety_{Beep}$	11	1
$Safety_{Voice}$	12	0
$Comfort_{Beep}$	5	7
$Comfort_{Voice}$	7	5

Figure 6: Results of the question if an override system is useful, ranging from 0 (not useful) to 5 (highly useful) and counted yes or no answers of the question if the participant want to have such function in his car.

All participants that experienced the spoken feedback in the safety scenario answered with "yes", because they liked the idea of increased safety and one participant stated it would be "*like a guardian angel driving with you*" (P9). The participants that tested the feedback with a beeping sound had similar opinions, except for P7, who thought to have the situation under control for himself and the "*beeping gave me the impression of having done something wrong*" (P7). The feedback we got from the participants after the non-safety related overrides was more diverse. The answers ranged from "*I like it, because I am often losing my way while driving*" (P10) to "*No, because you have the feeling of being in an emergency situation*" (P5). Participant 18 raised concerns that "*with that function I might lose my driving skills, because I always rely on it*".

We also asked the participants how they liked the feedback method of the system. The answers were categorized into the three categories "liked it", "liked it with idea for improvement" and "didn't like it". In both safety-related scenario groups, the participants generally liked the feedback, but three of the participants that experienced the beeping sound wished for a spoken cue (P13, P15, P17). The results from the comfort scenario were more diverse: eight of the twelve participants that experienced the interaction with only the beeping sound stated that this was not a good feedback method, because they did not get a concrete hint what the system was doing. In the group with spoken feedback only two of the twelve participants were unhappy with the feedback method. Participant 10 even questioned if the explanation is required at all and suggested the system to just state "*everything is fine – I take over now!*" (P10). 23 of the 24 participants wanted to get a warning in advance before the system is taking over control. Participant 1 stated that a notification of a (possible) takeover situation could be annoying for the driver.

Results: Video Analysis

To answer Q4 we had to observe the participants during the system interaction. For each participant four videos were recorded, two showing the upper part of the body to examine reactions and steering wheel interaction and two showing the participants' feet and gas and brake pedals. One set of videos was recorded while driving route 1 and the other set while driving route 2. For analysis, the behavior of the participants shortly before, during and after the override situations was of interest. The categorization was developed inductively during the analysis of the videos. The 12 subjects that were part of the safety override group with the beeping sound feedback had no distinct reaction during the system intervention. Some of them said something like "*oh*" or "*oops*". The pedal camera showed that most participants were pressing the gas pedal continuously during the system-initiated maneuver. The same behavior concerning the pedals was observed in the safety override group with spoken feedback, with the difference that some people stopped to press the gas pedal as soon as the spoken explanation was played. The same people's reaction to the explanation was very interesting: Afterwards these subjects waited longer than the others did, questioning if something would happen. One participant asked: "*may I now drive again?*" (P10). In both groups of the comfort-function, beeping sound feedback and spoken feedback, people tried to counteract the system's actions during the override, as well through counter steering and through braking or pressing the gas pedal. The difference between the two groups was in the behavior shortly after the system intervention. While the participants of the group with the beeping sound looked confused or frightened, the participants of the group with the voice feedback were all smiling or laughing. In general, the spoken explanation encouraged interaction of the participants with the system. They responded to the explanation with responses like "*ok*", "*thank you*", "*if you say so.*.", one participant thought to have the situation under control by himself and when the automation stated that it had to override him to prevent an accident he responded: “*don’t talk nonsense! You’re lying!*”(P8).

Results: Ethical Questionnaire Feedback

The IDE prepared a questionnaire, which was given to the participants after the driving situations. This was done to gather their feedback after experiencing a situation where their own autonomy to make decisions was delimitated by a machine. In the following paragraph, the questions and answers have been translated into English by the authors as closely as possible to the German original. The first question was: ”Is a vehicle already autonomous if it has non-overridable driving capabilities? What does a machine make autonomous in your opinion?” This question deliberately picks up the medial discourse where often autonomous driving is used as generic term [21], although this is not the correct term if the car is not capable of driving in SAE level 5. The answers to this question were meant to give an overview of the differences made between automation and autonomy, especially in an environment with high affinity for technology [12]. Many of the participants showed incertitude using the term “autonomy”. Very different explanations and opinions regarding concepts for autonomy in human-computer interaction were received. It seems to be difficult for the participants to differentiate between the term “autonomy” in partly-, highly- and fully-automated technologies. Nine of the participants classified the tested system as autonomous and stated that systems like parking assist, lane keeping assist or “*information on tire pressure*” (P9) are also autonomous. 15 participants classified the tested system to not being autonomous. We received several contradictory statements that indicate an existing confusion: ”*As soon as the driver can turn the vehicle on and off, it is not an autonomous car. But if the car takes away my emergency reaction already in minor situations, it is too autonomous in my opinion*” (P23). Another answer received was “*Autonomy in my opinion is if it is comfortable for me*” (P11). Participant 4 attributed human characteristics to our system: “*Yes, it is autonomous, respectively stubborn! By not being overridable, the machine gets its own will and becomes human. But that is in principle uncomfortable*“. This is consistent to the reactions of the participants to the spoken feedback described in the previous paragraph. A possible explanation could be a felt loss of control, which is expressed by imputing autonomy to the system. Losing control is perceived through the loss of own autonomy, which is then transferred to an increased autonomy of the system: “*An autonomous machine is capable of taking over control*” (P5). In this context participant 12 wrote: “*One must want to give up control and be able to do so, the trust in the system is missing and skepticism prevails*”. The answers we got could also indicate that one does not want to be responsible once “*the main responsibility*” (P7) has been transferred to the system.

The participants were also asked about trust in the system, to find out if they would give the system full responsibilities in critical situations: “What is a trustworthy system? Is there any difference between you trust in a human and your trust in technology?” Apparently, on the one hand, a distinct skepticism is present, but on the other hand, a willingness to handover control and responsibility to the system exists – provided that technology is 100% reliable. Eight participants preferred to hand over control to the car in critical situations or even suggested “*to shift the responsibility on to the car – if I want to*”(P8). While ten participants were undecided and mentioned a situational decision, six participants wanted to drive without an automation. Again, several inconsistent statements were received: “*As I do not like to let another person drive and I like to drive myself and I only trust in myself, I would trust the automation at very narrow roads in the mountains*” (P9). The answers indicate that the idea of an equal partnership between the human and the automation is rather to rely on the system and to hand over sovereignty to it: “*In my opinion the car should make the decision what needs to be done and therefore stand above my orders. Provided that these decisions are correct*” (P15). Technology even receives trust in advance: “*A trustworthy system needs to be reliable and its actions need to be comprehensible. If that is the case I trust the system more than I trust other people*” (P15). “*In principle I would rather trust technology than other people, because technology is predictable and people are not*” (P10). The participants also stated their concerns that their trust in the system could be easily shaken. The consequences of the loss of trust are expressed very clearly as one “*could never again trust technology*” (P9) or one “*would abstain from this technology in the future*” (P12). Only two participants stated that they would use such technology again after an erratic behavior and loss of trust. 14 participants stated that they would not use this kind of technology afterwards and eight did not give a clear statement regarding the use after an erratic behavior.

DISCUSSION

Considering the positive attitude of our participants regarding a function that can take over the driving task by itself if the driver is making a mistake, we can answer Q1 (*Are drivers open to a system capable of overriding them and do they want to have one in their car?*) with a clear “yes”. We therefore propose further research in this particular field. A differentiation has to be made if the function works like a guardian angel, intervening in hazardous conditions or if the function also takes over control in non-critical situations (Q2: *Do the drivers only allow an action of the system if the situation is critical or also in uncritical situations?*). While the first possibility was accepted by almost all our participants the second one yielded mixed reactions. A possible approach would be to make the latter a comfort function that can be switched off and on by the driver. Regarding Q3 (*How does a “guardian angel” need to communicate its intervention?*) our research shows that a spoken feedback facilitates more appreciation than a simple beeping sound. It has yet to be researched how a visual or haptic warning would improve the reaction of the users. All but one participant explicitly wanted to get a warning in advance, which was deliberately omitted in the study to focus on the overwriting situation. It has to be examined if there is time for a warning in advance as a possible overwrite situation may not be recognizable way ahead. A warning in advance could become annoying very fast if it was in situations where the driver can detect the danger on his or her own and react accordingly. Situations where the car overwrites the driver should happen rather rarely and therefore no negative effects of the spoken feedback as described in [19] should occur. Another open problem is the needed handover from the automation back to the driver after an intervention. Participants felt unclear whether or not they could take back control once the automation took over. The insights we gained from the video analysis will help us to determine future strategies, as we now know how people might respond to certain system actions (Q4: *How do drivers behave while the system is actively overriding them?*). As this paper focuses on the driver-vehicle-interaction, we also did not address the situational recognition of a hazardous situation and the deciding process if an intervention from the automation could resolve the imminent danger. It might be helpful for future research to determine the reason why an override is needed. Is it because of a failure of perception or a failure of proper steering by the driver? Was the situation created because of other road users or maybe even on purpose by the driver?

The tested "guardian angel"-function is located in a border area between automation and autonomy, which is the reason Q5 (*Do people already have a consistent mental concept of the role model of a driver of an automated vehicle*?) was asked. It is difficult for the users to classify such a function, as the authority of control of the human is being revoked for a short time. Together with the inconsistent classification of autonomy, the question arises if the changed concept of responsibilities for (highly) automated driving is already understood and accepted by the users. This is going to be of particular importance in the design of the human-machine-interaction. Our results show that, especially from an ethical point of view, the users need to have a good understanding of their own responsibilities regarding autonomous systems and system borders. Highly automated systems do not work autonomously [2] and functions like parking assist or lane keeping assistance cannot be used without human guidance. Monitoring a system needs knowledge in one's own authority of control and sovereignty. Research has to survey if users are aware of this connection and their own responsibility. Hopes and expectations regarding automotive systems are high and bound to a clear requirement: Systems are not allowed to make mistakes.

Combining the qualitative and the ethical questionnaire another important point that factors into the answer of Q1 can be witnessed: Our questionnaires showed a distinct unsteady opinion about the human-machine-interaction in the context of automated driving. 23 of our 24 participants clearly wanted to have a safety-related guardian angel in their car, like the one proposed to them during the study. When we asked them in a more abstract way in the ethical questionnaire, only 8 participants were willing to hand over control to the system in critical situations.

SUMMARY

For future development, we need sensitization, especially for the designers and developers. Our research showed that many concepts regarding responsibilities during automated driving are not clear to the users. The current concepts of automation might not be easily conclusive. There are abstract concepts, like the autonomy of the car, that are complex and need to be simplified. Critical takeovers and takeover situations are not self-explanatory concerning when the human needs to be ready and when not; in contrast our study showed that there are contradictory views of the users. From an ethical perspective, we need to raise awareness for the designers and engineers how to name a certain system because of the attributes the users tend to give it. Designing a system in the border zones between SAE levels 3 to 5 needs to be done with caution, with regard to the perceived system capabilities.

Clearly, a "guardian angel" can save lives in future road traffic, if the drivers of future cars are willing to have such an assistive system on board. This depends heavily on the design of such a system. A guardian angel is an entity that can sometime save lives, if it has the possibility to do so; It is not an entity that takes full responsibility for safe driving.

ACKNOWLEDGEMENTS

This Work was funded by the German Federal Ministry for Education and Research (ID: 16SV7622K). The authors are responsible for the content of this publication.

REFERENCES

1. Bainbridge, Lisanne. 1982. "Ironies of automation". *Analysis, Design and Evaluation of Man–Machine Systems 1982*: 129-135.
2. Both, Göde; Weber, Jutta. 2013. "Hands-Free Driving? Automatisiertes Fahren und Mensch-Maschine Interaktion". *Robotik im Kontext von Recht und Moral*, Volume 3: 171-189.Nomos Verlagsgesellschaft mbH & Co. KG
3. Bruemmer, David J., Donald D. Dudenhoeffer, and Julie L. Marble. 2002. "Dynamic-Autonomy for Urban Search and Rescue". *AAAI Mobile Robot Competition.*
4. Desai, Munjal, and Holly A. Yanco. 2005. "Blending human and robot inputs for sliding scale autonomy". *Robot and Human Interactive Communication*. IEEE International Workshop on IEEE 2005.
5. Filippidou, Despina. 1998. "Designing with scenarios. A critical review of current research and practice". *Requirements Eng* 3 (1): 1–22. DOI: 10.1007/BF02802918.
6. Grimm, Petra; Müller, Michael. 2016. „Narrative Medienforschung. Einführung in Methodik und Anwendung". UVK Verlagsgesellschaft mbH.

7. Grundmann, Thomas; Horvath, Joachim; Kipper, Jens. 2014. „Die experimentelle Philosophie in der Diskussion“. Suhrkamp.

8. Grundmann, Thomas; Horvath, Joachim; Kipper, Jens. 2014. „Die Experimentelle Philosophie in der Diskussion. Eine Einleitung“. *Die experimentelle Philosophie in der Diskussion*: 9-55. Suhrkamp.

9. S.G. Hart, L.E. Staveland. 1988. “Development of NASA-TLX (Task Load Index): Results of Empirical and theoretical research”. *Advances in psychology 52:* 139-183.

10. S.G:Hart. 2006. “NASA-task load index (NASA_TLX): 20 years later”. *Proceedings of the human factors and ergonomics society annual meeting, Vol 50:* 904-908. Sage Publications

11. Hassenzahl, Marc, Michael Burmester, Koller, Franz. 2003. „AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität“. *Mensch & Computer:*187-196. Vieweg + Teubner Verlag

12. Hilgendorf, Eric. 2015. „Teilautonome Fahrzeuge: Verfassungsrechtliche Vorgaben und rechtspolitische Herausforderungen“. *Rechtliche Aspekte automatisierter Fahrzeuge. Beiträge zur 2. Würzburger Tagung zum Technikrecht*:15-32.

13. Hwang, Wonil, and Gavriel Salvendy. 2010. “Number of people required for usability evaluation: the 10±2 rule”. *Communications of the ACM* 53.5: 130-133

14. Jacobson, Ivar; Spence, Ian; Kerr, Brian. 2016. “Use-Case 2.0. The Hub of Software Development”. *ACM Queue, Vol. 14 (1)*:94-123.

15. J. Koo, J. Kwac, W. Ju , M. Steinert, L. Leifer and C. Nass. 2015. “Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance" *International Journal on Interactive Design and Manufacturing:* 269-275.

16. Löper, Christian, Johann Kelsch, Flemisch, Frank Ole. 2008. "Kooperative, manöverbasierte Automation und Arbitrierung als Bausteine für hochautomatisiertes Fahren".

17. S. Maurer, E. Rukzio and R. Erbach. 2017. "Challenges for Creating Driver Overriding Mechanisms". *Adjunct Proceedings of the 9th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications*

18. Mukerji, Nikil; Knobe, Joshua. 2016. Einführung in die experimentelle Philosophie. Wilhelm Fink.

19. F. Nothdurft, S. Ultes and W. Minker. 2015. "Finding appropriate interaction strategies for proactive dialogue systems - an open quest". *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication*.

20. Quinn, Regina Ammicht; Nagenborg, Michael; Rampp, Benjamin; Wolkenstein, Andreas F.X. 2014. „Ethik und Sicherheitstechnik. Eine Handreichung“. *Sicherheitsethik*: 277-296. Springer VS

21. Reek, Felix. 2018. „Der Mensch ist noch zu kindisch für autonomes Fahren“. *Süddeutsche Zeitung, 25. Januar 2018*. Retrieved January 29, 2018 from http://www.sueddeutsche.de/auto/tesla-der-mensch-ist-noch-zu-kindisch-fuer-autonomes-fahren-1.3839832.

22. Sheridan, T., Verplank W. 1978. “Human and Computer Control of Undersea Teleoperators”.

23. Simon, Judith. 2016. “Values in Design“. *Handbuch Medien- und Informationsethik*: 357-364. J.B. Metzler.

24. Stanton, Neville A., Salmon, Paul M. 2009. "Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems." *Safety Science 47.2*: 227-237

25. Winner, Hermann, Hakuli, Stephan. 2006. "Conduct-by-wire–following a new paradigm for driving into the future." *Proceedings of FISITA world automotive congress. Vol. 22*.

26. Winner, Hermann, Hakuli, Stephan, Wolf, Gabriele. 2011. „Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort“. Springer.

27. Young, M. S., Stanton, N. A., Harris, D. 2007. “Driving automation: learning from aviation about design philosophies”. *International Journal of Vehicle Design*, *45*(3): 323-338.

28. Lee, John D., Joshua D. Hoffman, and Elizabeth Hayes. 2004. "Collision warning design to mitigate driver distraction." *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*.

29. Knecht, Chiara, et al. 2014 "Vertrauen in die Automatisierung, fehlende Situation Awareness & Fertigkeitsverlust durch automatisierte Systeme–Eine subjektive Einschätzung aus der Linienpilotenperspektive." *56. Fachausschusssitzung Anthropotechnik der DGLR*.

30. Bosch Mobility Solutions. 2018. „ New service for drivers: Bosch lets cars find parking spaces themselves “. Retrieved May 4, 2018 from https://www.bosch-mobility-solutions.com/en/highlights/connected-mobility/community-based-parking/.

31. Bundesministerium für Verkehr und digitale Infrastruktur. 2017. „Ethik-Kommission Automatisiertes und Vernetztes Fahren. Bericht Juni 2017“. Retrieved July 17, 2017 from https://www.bmvi.de/SharedDocs/DE/Anlage/Presse/084-dobrindt-bericht-der-ethik-kommission.pdf?__blob=publicationFile

32. Kaplan, Jeremy. 2018. “Here’s every company developing self-driving car tech at CES 2018”.

Retrieved May 6, 2018 from https://www.digitaltrends.com/cars/every-company-developing-self-driving-car-tech-ces-2018/

33. “Institut für Digitale Ethik”. Retrieved April 25, 2018 from https://www.digitale-ethik.de/.

34. „KoFFI - Kooperative Fahrer-Fahrzeug-Interaktion“. Retrieved April 25, 2018 from https://www.technik-zum-menschen-bringen.de/projekte/koffi

35. SAE. “Automated Driving”. Retrieved June 13, 2017 from http://www.sae.org/misc/pdfs/automated_driving.pdf

36. WIVW. “Fahrsimulation und SILAB”. Retrieved May 6, 2018 from https://wivw.de/de/silab